



## Practice of Epidemiology

# A Vaccine Study Design Selection Framework for the Postlicensure Rapid Immunization Safety Monitoring Program

Meghan A. Baker\*, Tracy A. Lieu, Lingling Li, Wei Hua, Yandong Qiang, Alison Tse Kawai, Bruce H. Fireman, David B. Martin, and Michael D. Nguyen

\* Correspondence to Dr. Meghan A. Baker, Department of Population Medicine, Harvard Pilgrim Health Care Institute and Harvard Medical School, 133 Brookline Avenue, 6th floor, Boston, MA 02215 (e-mail: meghan\_baker@harvardpilgrim.org).

Initially submitted December 26, 2013; accepted for publication October 16, 2014.

The Postlicensure Rapid Immunization Safety Monitoring Program, the vaccination safety monitoring component of the US Food and Drug Administration's Mini-Sentinel project, is currently the largest cohort in the US general population for vaccine safety surveillance. We developed a study design selection framework to provide a roadmap and description of methods that may be utilized to evaluate potential associations between vaccines and health outcomes of interest in the Postlicensure Rapid Immunization Safety Monitoring Program and other systems using administrative data. The strengths and weaknesses of designs for vaccine safety monitoring, including the cohort design, the case-centered design, the risk interval design, the case-control design, the self-controlled risk interval design, the self-controlled case series method, and the case-crossover design, are described and summarized in tabular form. A structured decision table is provided to aid in planning of future vaccine safety monitoring activities, and the data components comprising the structured decision table are delineated. The study design selection framework provides a starting point for planning vaccine safety evaluations using claims-based data sources.

immunization; methods; Mini-Sentinel; PRISM; safety; surveillance; vaccine

Abbreviations: HOI, health outcome of interest; PRISM, Postlicensure Rapid Immunization Safety Monitoring; SCCS, self-controlled case series; SCRI, self-controlled risk interval.

The Postlicensure Rapid Immunization Safety Monitoring (PRISM) Program is the immunization safety monitoring component of the Food and Drug Administration's Sentinel Initiative, a program created in response to a congressional mandate to develop a national postmarket risk identification and evaluation system for the Food and Drug Administration-approved medical products using electronic health-care data (1). PRISM is a distributed data network with claims data from 4 national health insurers and vaccine data from 8 immunization registries from state and city health departments; it is currently the largest geographically and demographically diverse population-based cohort in the United States for active vaccine safety surveillance (1, 2). In PRISM, data partners maintain direct control over their data, and large-scale epidemiologic assessments are made possible by analytical programs that are distributed to all participating sites that run

against a common data model to enable consistent application of analyses across multiple data sources. This achieves the requisite large sample sizes necessary to study the subgroups and rare outcomes routinely evaluated in vaccine safety, while avoiding the pooling of large amounts of highly detailed individual-level data in a central location that would otherwise raise privacy concerns. Yet the advantages gained from using multisite claims-based data networks to evaluate vaccine safety also involve trade-offs that impact study design selection.

We created a study design selection framework for the evaluation of vaccine safety concerns in PRISM and other databases using administrative data. These safety concerns may arise during product development or after licensure; they may also be theoretical risks based upon commonalities among products of the same class or containing similar product

components (3). This paper provides an overview of epidemiologic methods that have been implemented in distributed database environments to assess vaccine safety, and it builds upon other work in the Food and Drug Administration's Sentinel Initiative and the Centers for Disease Control and Prevention's Vaccine Safety Datalink (3–7). We begin with a general discussion of the unique features of vaccine safety assessments in a distributed network that drive study design selection and follow with a description of 7 methods in terms of the analytical populations assembled, the data compared, and their overall strengths and weaknesses in the PRISM environment (Table 1). In the end, we provide a structured decision table to help guide Sentinel investigators in selecting study designs (Tables 2 and 3).

### FEATURES OF VACCINE SAFETY SURVEILLANCE IN A DISTRIBUTED CLAIMS-BASED DATA ENVIRONMENT THAT IMPACT STUDY DESIGN SELECTION

PRISM's longitudinal health-care database contains medical and pharmacy claims from Aetna, Inc. (Hartford, Connecticut), HealthCore, Inc. (Wilmington, Delaware), Humana, Inc. (Louisville, Kentucky), and Optum, Inc. (Eden Prairie, Minnesota), that is managed by a single coordinating center responsible for collaborating with Data Partners to transform their source data using a common data model, verifying its completeness and accuracy, and developing and testing the distributed analytical code. With claims data, exposure and outcome misclassification are important considerations in study design selection. Early experience with PRISM claims-based algorithms to identify vaccine exposure has found the positive predictive values to be high among cases with available charts, ranging from 94% to 100% for inactivated influenza, pneumococcal, and tetanus-containing childhood vaccines (8). Despite excellent confirmation rates, the high population level uptake of many vaccines, concerns regarding unmeasured differences between unvaccinated and vaccinated persons, and the difficulty in confirming the vaccination status for those who do not have a claim for immunization compel many vaccine safety study investigators to seek designs restricted to vaccinated persons to reduce bias.

In contrast, confirmation rates for health outcomes in PRISM have been more variable, compared with distributed networks comprising mostly electronic medical records such as the Vaccine Safety Datalink. For example, the positive predictive values for intussusception and venous thromboembolism (preliminary data only) are substantially lower, even after accounting for code selection differences, whereas the positive predictive value for febrile seizures was nearly equivalent to electronic medical record-based networks (8–12). Given the uncertain accuracy of many claims-based algorithms in PRISM, chart review may continue to be necessary, favoring designs that minimize this burden or implicitly control for confounding that would otherwise require more detailed covariate data to be collected to enable adjustment.

The medical record validation process for claims-based databases is more time intensive and costly, because charts do not reside in an integrated electronic information system but are spread across different health systems and care settings. Investigators request specific charts and chart components that must

be retrieved from health-care providers, a process that can take months. Furthermore, additional information to validate exposures and covariates may require data linkage to complementary data sources, such as immunization and birth certificate registries, which adds to the time and cost burden.

Finally, distributed databases protect privacy by using aggregate data at the cost of analytical flexibility. Avoiding the pooling of large amounts of patient-level information across sites is an important driver in study design selection. This data milieu favors designs using within-site matching of covariates, methods that rely on site-specific confounding adjustment scores, and distributed regression approaches so that sites transfer only summary statistics for model fitting (3–5, 13).

Multisite distributed databases enable large-scale epidemiologic assessments of vaccine safety with greater statistical power, precision, and speed (through more rapid patient accrual) than ever before and uncover safety information on new vaccines for public health intervention. However, the traditional study designs that power most single database studies generally must be adapted to leverage the advantages conferred by the multisite structure and to overcome important data constraints. This review provides an overview of several such design adaptations adopted by PRISM.

### COHORT DESIGN

The cohort design studies vaccinated persons and unvaccinated persons or persons vaccinated with a different vaccine over a period of time, and it compares the incidence of the health outcome of interest (HOI) across these groups. A defined population of individuals is identified and classified on the basis of their vaccination status. The design may be conducted prospectively or retrospectively, and the unexposed group or comparator group may be historical or concurrent with the exposure of interest. Persons or person-time, time units contributed by persons at risk, may serve as denominators for incidence, to be compared between the exposed and unexposed. The cohort study design is advantageous in a situation where an exposure is rare and multiple outcomes related to the same exposure are of interest in the same population (6, 14, 15). For vaccine safety, the cohort design may be difficult to implement if high vaccine coverage leaves few unvaccinated individuals available for comparison or when there are significant concerns regarding the comparability of unexposed to exposed populations because of socioeconomic status, race, underlying health conditions, or access to health care even after explicit adjustment for measured confounders (6, 16). Moreover, the burden of chart review is high as both unexposed and exposed groups need validation. Although a variety of analytical approaches are available, vaccine HOIs are typically examined as acute events or dichotomous outcomes. Logistic regression, Poisson regression, or Cox regression is used to estimate a multiplicative measure of the association of vaccination with the risk of the HOI, such as the relative risk or a closely related effect measure such as odds ratio (logistic regression), rate ratio (Poisson regression), or hazard ratio (Cox regression). The absolute risk and the attributable risk can then be estimated by combining the relative risk estimate with information about risk of the HOI in the comparison group (or a reference population).

Table 1. Summary Table

Study Design	Population	Example of Data Being Compared	Strengths	Weaknesses	Examples of Test Statistics or Regression Models
Cohort study	Vaccinated persons Unvaccinated persons/persons vaccinated with a control vaccine (often matched on potential confounders)	Incidence or incidence rates of those vaccinated vs. incidence or incidence rates of those unvaccinated	Standard design Easily implemented in PRISM because of the large amount of data available Matching controls for potential confounders Adjusts for seasonality; index date aligns risk period for vaccinees and controls	Confounding by indication, other unmeasured confounders Susceptible to misclassification of exposure Unvaccinated control population may be limited Biased incidence rate ratio if loss to follow-up is affected by both exposure and disease status	Linear and logistic regression Cox regression Poisson regression (conditional logistic regression or Poisson regression if matched analysis)
Case-centered analysis	Cases	Odds of vaccination during case window vs. odds of vaccination during control window; uses additional information on distribution of vaccination time	Adjusts for seasonality of vaccination and HOIs Facilitates data management and privacy in multisite studies with distributed data	Does not implicitly adjust for confounders other than seasonality, time, and age May lack power if cases occur early or late in study period, noninformative cases	Logistic regression with offset terms
Risk interval	Vaccinated persons	Incidence rates of exposed time periods vs. incidence rates of unexposed time periods	Time periods before and after vaccination used; ideal for acute self-limited events after vaccination Cases and noncases informative Less susceptible to misclassification of exposure	Confounding by indication, other unmeasured confounders Self-controlled only if proportion of exposed to unexposed time for those who develop the HOI is the same as proportion for those who do not	Poisson regression

Table continues

In distributed databases, matching or stratification may be used efficiently to ensure that the vaccinated and unvaccinated groups are comparable on specific characteristics, decreasing the potential for confounding from time-stable factors (e.g., sex and genetics) and also from time-varying confounders (e.g., seasonality) that change in the same way and at the same time for the vaccinees and their matches (Figure 1). Matching can address site heterogeneity as matching is typically performed within each Data Partner. It can also avoid pooling individual-level data on the entire cohort as pooling of data is needed only for the matched cohort. A 2-phase sampling design may also avoid pooling a large amount of individual-level data by using information on the whole cohort to identify the most informative subgroup on which to collect supplemental data and to facilitate more complete confounder adjustment (17, 18).

A cohort design was recently used by Shui et al. (12) to examine the risk of intussusception among US infants following the pentavalent rotavirus vaccine. In this study, the rate of intussusception in infants 4–34 weeks of age in the Vaccine Safety Datalink population who received the pentavalent rotavirus vaccine over a 4 year period was compared with the rate among infants who did not receive the pentavalent rotavirus vaccine but did receive other recommended vaccines during the same time period.

## CASE-CENTERED APPROACH

The case-centered approach was developed to address time-varying confounding in vaccine safety studies (19). This approach examines whether there is an association of the HOI with the vaccine by testing whether the case was more likely to be vaccinated during a risk interval prior to the HOI onset date compared with similar people (i.e., subjects in the same stratum defined by important confounders) from the general population or, equivalently, whether the vaccinee was more likely to develop the HOI in a prespecified risk window compared with similar individuals from the general population. The case-centered approach may either look backward in time from the date of the HOI as in the case-control design (Figure 2) or look forward from the vaccination date as in the cohort design.

Expectations about the probability that a case was recently vaccinated are derived from the case's stratum of the population. In a hypothetical example where there is concern about risk of an HOI within 1 week of influenza vaccination, one outcome occurs on November 1 when 20% of vaccinees are within 1 week of vaccination, and another event occurs on November 15 when 10% of vaccinees are within 1 week of vaccination. If vaccination does not increase risk, we would

Table 1. Continued

Study Design	Population	Example of Data Being Compared	Strengths	Weaknesses	Examples of Test Statistics or Regression Models
Case-control	Cases Noncases (often matched on potential confounders) Assumption: control group originated from the same hypothetical population that gave rise to the cases	Odds of vaccination among cases vs. odds of vaccination among the control group	Standard design Uses small data sample from entire group, cost-efficient Time-varying confounders may be controlled for by matching	Confounding by indication, other unmeasured confounders Selection bias Susceptible to misclassification of exposure Unvaccinated population may be limited Biased incidence rate ratio if loss to follow-up is affected by both exposure and disease status	Logistic regression (conditional logistic regression if matched analysis)
Self-controlled risk interval	Vaccinated persons, but only cases informative	Incidence rates of exposed time periods vs. incidence rates of self-matched unexposed time periods	Self-controlled, adjusts for time-invariant confounders Less susceptible to misclassification of exposure	Time-varying confounding (less susceptible than SCCS because of less variation in periods under observation) Reverse causality bias Only cases informative, reducing efficiency	Conditional Poisson regression
Self-controlled case series method	Cases	Incidence rates of exposed time periods vs incidence rates of self-matched unexposed time periods	Self-controlled, adjusts for time-invariant confounders Multiple occurrences of independent events within an individual can be assessed	Time-varying confounding Reverse causality bias	Conditional Poisson regression
Case-crossover analysis	Cases	Odds of vaccination during case window vs. odds of vaccination during control window	Self-controlled, adjusts for time-invariant confounders	Time-varying confounding Exposure trend bias	Conditional logistic regression Conditional Poisson regression

Abbreviations: HOI, health outcome of interest; PRISM, Postlicensure Rapid Immunization Safety Monitoring; SCCS, self-controlled case series.

calculate expectations that 0.02, 0.72, or 0.26 would be the probability that both, neither, or 1 of the events, respectively, occurred during the risk interval. (The calculations are as follows:  $0.2 \times 0.1$  yields a 0.02 chance of both events occurring inside the interval,  $0.8 \times 0.9$  yields a 0.72 chance of both occurring outside the interval, and the remaining possibilities amount to a 0.26 chance that just 1 of the events was in the interval). Thus, in this example where there were only 2 outcome events in vaccinees and both events occurred during the risk interval, the null hypothesis is rejected (that vaccination is unrelated to risk), with an “exact” one-sided  $P = 0.02$ .

The association of the vaccine with risk may be estimated by fitting a simple logistic regression model to a summarized data set including only 1 record per risk set. The outcome variable indicates whether the outcome event occurred in the case’s risk interval, and the key predictor variable is the proportion of the risk set who were in their risk interval on the date of the case’s outcome event. By anchoring the risk sets to calendar dates, time-dependent confounding, such as confounding by seasonality in outcome incidence and vaccine delivery, is minimized. To further adjust for potential confounders and to facilitate evaluation of effect modification, risk sets can be restricted to vaccinees who are similar with

respect to known factors that might be confounders or effect modifiers, such as age group or Data Partner.

The results of the logistic regression (with only 1 record per outcome event) include an estimate of the odds ratio and a corresponding confidence interval assessing how much the risk is elevated during the risk interval. These results are identical to the hazard ratio, confidence interval, and hypothesis test that would be obtained by a cohort study using stratified Cox regression fit to individual-level data with the same risk sets on the same calendar timeline (20). A motivation for the case-centered specification of the regression is that it is an intuitive way to focus on change in risk on a time-since-vaccination timeline while adjusting carefully for seasonality and other potential confounders on a calendar timeline. It also facilitates more direct examination of heterogeneity in the relative risk across age groups, Data Partners, and other subgroups. An important advantage is that the case-centered approach minimizes privacy concerns, because only aggregated data on each risk set are required for the overall analysis. Thus, patient-level data can remain in distributed databases and need not be pooled for analysis.

Limitations of the case-centered approach include the requirement that adjustment for patient-level covariates be done by stratification (rather than by adding covariates to the model).

**Table 2.** Structured Decision Table

HOI Onset	Duration-of-Exposure Risk Interval	Strength of Confounding		Design Options
		Time-Invariant Factors	Time-Variant Factors (e.g., Seasonality or Age)	
<i>Misclassification Negligible</i>				
Well defined	Short	Negligible	Negligible	Cohort, SCCS, SCRI, case-crossover, risk interval Case-centered analysis, cohort SCCS, SCRI, case-crossover Case-centered analysis <sup>a</sup>
		Not negligible	Not negligible Negligible Not negligible	
Insidious	Long	Negligible	Negligible	Cohort Case-centered analysis, cohort Matched cohort Case-centered analysis, matched cohort
		Not negligible	Not negligible Negligible Not negligible	
<i>Misclassification Not Negligible</i>				
Well defined	Short	Negligible	Negligible	SCCS, SCRI, case-crossover, risk interval Case-centered analysis SCCS, SCRI, case-crossover Case-centered analysis <sup>a</sup>
		Not negligible	Not negligible Negligible Not negligible	
Insidious	Long	Negligible	Negligible	Cohort Case-centered analysis, cohort Matched cohort Case-centered analysis, matched cohort
		Not negligible	Not negligible Negligible Not negligible	

Abbreviations: HOI, health outcome of interest; SCCS, self-controlled case series; SCRI, self-controlled risk interval.

<sup>a</sup> SCRI and case-time-control designs both have some capacity to adjust for time-varying confounding. Depending on the bias and the availability of information to adjust for time-varying confounding, these designs may be preferable.

The marking of time as days rather than weeks or months may entail such fine stratification that some health outcomes become uninformative and power is diminished. The case-centered

approach also remains vulnerable to confounders that may vary within strata, if differences between recent and less-recent vaccinees are associated with risk of the HOI.

**Table 3.** Components of the Structured Decision Table

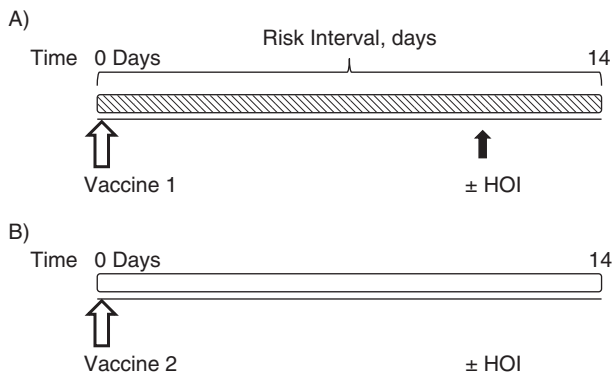
Study Designs	Cohort Study	Case-Centered Analysis	Risk Interval	Case-Control	Self-Controlled Risk Interval	Self-Controlled Case Series Method	Case-Crossover Analysis
Data characterization (proportion of patients on whom data are required)	Maximal	Moderate	Moderate	Moderate	Moderate	Minimal	Minimal
Misclassification bias if vaccine exposure not recorded in data set?	Yes	Yes (no if study population restricted to vaccinated persons)	No	Yes	No	Yes <sup>a</sup> (no if study population restricted to vaccinated persons)	Yes <sup>a</sup> (no if study population restricted to vaccinated persons)
Is design suitable for HOI with a late onset?	Yes	No	No	Yes	No	No	No
Is design suitable for HOI with insidious onset?	Yes	No	No	Yes	No	No	No
Is control for time-invariant factors implicit?	No	No	No	No	Yes	Yes	Yes
Is control for time-variant factors implicit?	No	Yes	No	No	No	No	No
Examples of methods to adjust for time <sup>b</sup>	Matching, modeling		Modeling	Matching, modeling	Modeling, offset term	Modeling	Case time control design, modeling

Abbreviation: HOI, health outcome of interest.

<sup>a</sup> Misclassification of vaccinated persons as unvaccinated persons may bias the estimate in a multivariate analysis.

<sup>b</sup> Not a comprehensive list.





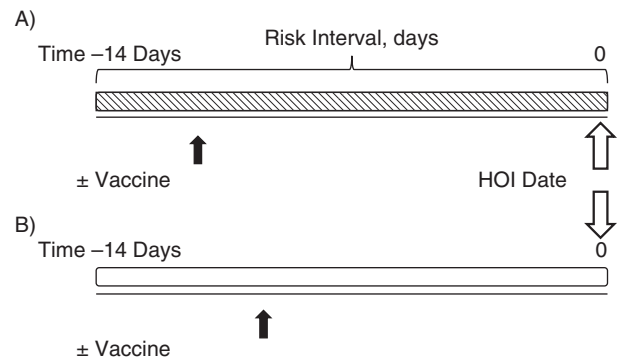
**Figure 1.** Matched cohort design with alternate vaccine comparator. A) Individuals vaccinated with the vaccine of interest (vaccine 1); B) individuals vaccinated with an alternate vaccine (vaccine 2). Persons vaccinated with the vaccine of interest or a comparator vaccine are followed for a distinct period of time, and the incidence of the health outcome of interest (HOI) in the 2 groups is compared. The striped bar represents the window of time at risk for the HOI after the vaccine of interest.

Rowhani-Rahbar et al. (20) described the case-centered approach in a study assessing the risk of Bell's palsy after the trivalent influenza vaccine, the hepatitis B virus vaccine, and any vaccine in children aged 18 years or younger. This analytical approach was used to control for confounding by age and seasonality by assessing the expected odds of vaccination during the risk interval based on each case of Bell's palsy. The authors used logistic regression with a data set limited to Bell's palsy cases. However, for each day that a case of Bell's palsy occurred, information on all similar people at risk for an immunization within the risk interval was included in the model as an offset term.

## RISK INTERVAL DESIGN

The primary strength of the risk interval design is that only vaccinated subjects are analyzed, minimizing bias due to the many unmeasured ways that vaccinees may differ from the unvaccinated, a limitation of the cohort design. Incidence of the HOI during risk periods is compared with incidence in control time periods that correspond to baseline risk, the expected risk had the vaccinee been unvaccinated or given a comparator vaccine (Figure 3). All vaccinated subjects (cases and non-cases) contribute to the risk estimate, and person-time is pooled for the risk interval and the control interval. This design is ideal for HOIs that are acute and brief (6).

Both pre- and postvaccination control intervals can be used. However, a prevaccination control interval may appear to be relatively low risk because of a healthy vaccinee effect (6, 16). To address the healthy vaccinee effect, we can exclude the time period immediately prior to vaccination from the analysis. Furthermore, because the distribution of vaccine risk over time is usually unknown, the vaccine risk may carry over into the control period, leading to an underestimate of the relative risk. A wash-out period after the risk interval



**Figure 2.** Case-centered design. A) Case; B) risk set. The rates of vaccination inside and outside the risk intervals are compared among subjects who experience at least 1 health outcome of interest (HOI). For each HOI case date, the vaccination status and timing of all individuals with the same covariate profile who were at risk for the HOI on that date are also analyzed as an offset term, allowing for adjustment of time-varying confounding.

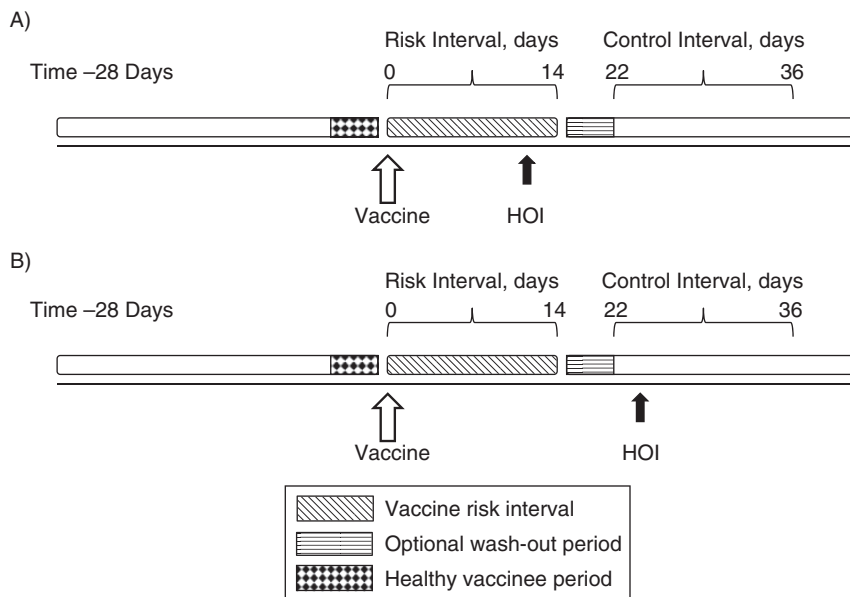
may be inserted before the control interval to minimize the residual risk attributable to the vaccine.

France et al. (21) used the risk interval design to assess the risk of immune thrombocytopenic purpura after measles, mumps, and rubella immunization. Only vaccinated individuals were included in the analysis, and follow-up time included 365 days before and after vaccination, excluding the 6 weeks prior to vaccination to avoid the healthy vaccinee effect. The exposed window was well defined, 42 days after measles, mumps, and rubella vaccination, and the incidence of the HOI was compared between the exposed and unexposed time periods.

## CASE-CONTROL DESIGN

Vaccination status is compared between subjects who experience an HOI (i.e., cases) and a control group who do not experience the HOI, who were chosen from the same population (Figure 4). The analysis provides an estimate of the odds ratio, an approximation of the relative risk when the incidence of the outcome is low in both the vaccinated and unvaccinated populations. Within-site matching of cases and controls can be used efficiently in distributed settings to control for potential confounding (6, 22, 23). Matching avoids pooling individual data for a large population, because data are needed only for cases and their matched controls. The study design is economical for HOIs that occur rarely, but identifying an appropriate control group may be the limiting factor.

Irving et al. (24) used the case-control design to examine the association between spontaneous abortion and maternal influenza vaccination. Cases of spontaneous abortion were matched by health-care organization and last menstrual period to controls with a livebirth. The odds of maternal vaccination during a risk interval corresponding to the 28 days prior to the date of spontaneous abortion of the matched pair were compared in cases and controls. The case-control design in this setting

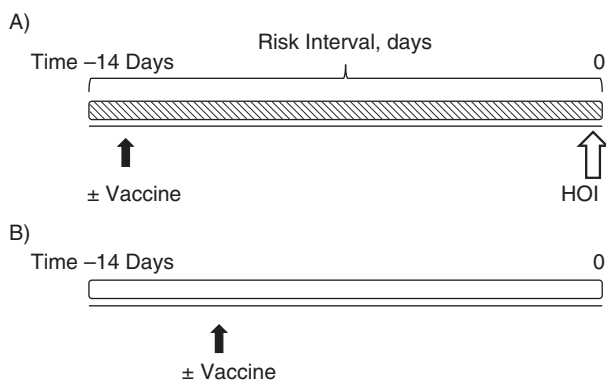


**Figure 3.** Risk interval design and self-controlled risk interval design. A) Subject 1; B) subject 2. Vaccinated individuals comprise the study population, and incidence rates for risk and nonrisk time periods are compared. In example 1, the vaccinated individual experiences a health outcome of interest (HOI) that falls within the risk interval. In example 2, the vaccinated individual experiences an HOI that falls within the control interval.

was resource efficient, because only the cases and controls required chart review. Matching on last menstrual period and accounting for the matching via conditional logistic regression inherently adjusted for gestational age and seasonality.

### SELF-CONTROLLED RISK INTERVAL DESIGN AND SELF-CONTROLLED CASE SERIES METHOD

The self-controlled risk interval (SCRI) design builds upon the risk interval design by self-matching the risk and

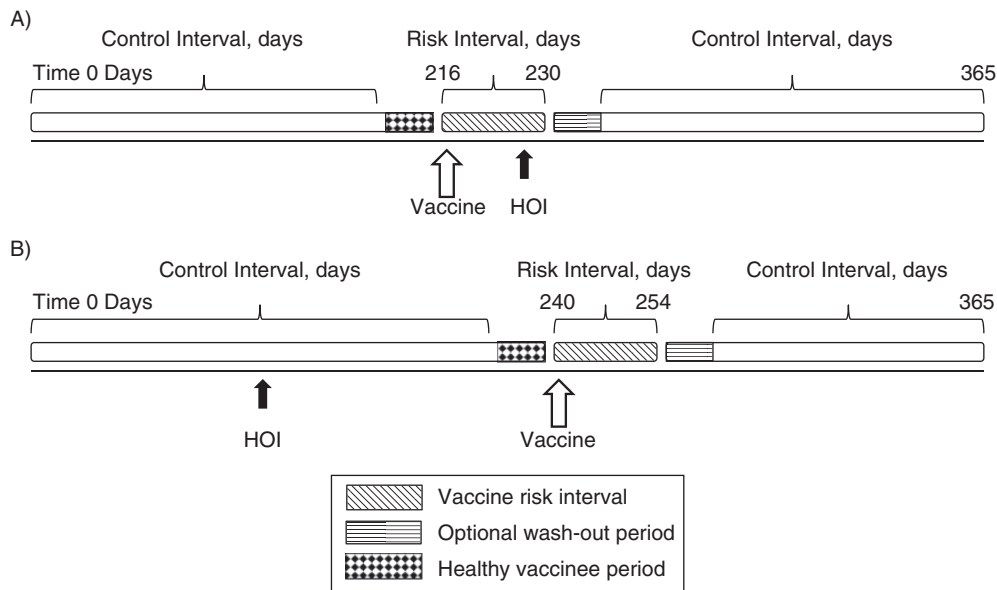


**Figure 4.** Case-control design. A) Cases; B) controls. Vaccination status is compared between subjects who experience a health outcome of interest (HOI) and a control group who do not experience an HOI. The striped bar represents the window of time prior to the HOI when a patient is at risk for vaccination.

control intervals, thereby implicitly controlling for time-stable confounders (Figure 3). This design compares the risk of the HOI in the risk interval with that in the control interval. Each individual contributes 1 or more control intervals before exposure and/or after exposure. Although the study population includes all vaccinees, only vaccinated cases may contribute to the risk estimation in the analysis.

The self-controlled case series (SCCS) method is similar to the SCRI in that the risk and control intervals are self-matched, but the study base is traditionally restricted to all cases (both vaccinated and unvaccinated) (Figure 5) (25–28). Each vaccinated case contributes both exposed and unexposed person-time, and the incidence of the HOI in the exposed person-time is compared with the incidence in the unexposed person-time. In this design, unexposed time may occur before or after vaccination as the standard SCCS method is bidirectional (21).

Both the SCRI and SCCS designs are self-controlled and therefore adjust for all time-stable confounders, including site heterogeneity, but are still vulnerable to time-varying confounders, including age and seasonality. One may control for time-varying confounding by either 1) using a nonparametric approach, in which the time periods are divided into subintervals on the basis of the values of the time-varying confounders, and additional terms for these subintervals are added in the conditional Poisson regression model (27), or 2) adjusting the expected incidence of the HOI over time according to the incidence in either the underlying population or an external population. In both designs, the conditional Poisson regression provides a direct estimation of the relative risk relevant to a population from which the cases were derived.



**Figure 5.** Self-controlled case series. A) Subject 1; B) subject 2. Subjects who experience at least 1 health outcome of interest (HOI) during a predefined study period are included. Vaccinated subjects contribute both exposed and unexposed person-time based on the timing of the vaccination and the predetermined risk interval. Unexposed control time may occur before or after vaccination and is often longer than in the self-controlled risk interval study. In example 1, the vaccinated individual experiences an HOI that falls within the risk interval. In example 2, the vaccinated individual experiences an HOI that falls within the unexposed interval.

If a prevaccination period is used as a control interval, an assumption of the SCCS and SCRI is that the occurrence of the HOI must not alter the probability of subsequent exposure to the vaccine. Violation of this assumption because of vaccine contraindication causes bias in risk estimation (29). Both self-controlled designs also require well-defined and transient risk intervals after exposure. Selecting a risk interval that is too wide or too short will bias the risk estimate relative to the true risk window.

The distinguishing feature of the 2 designs is the observation period. In the SCRI design, the vaccination date is typically used as the index date to define the risk and control intervals, and person-time is measured in time since vaccination. In contrast, in the SCCS design, a preselected observation period is created independent of the vaccination date, and all cases occurring within the observation period are identified. Similar to SCRI, the risk interval is predefined; however, the control interval is the observation period minus the risk interval. If the case is unvaccinated, the entire observation period is considered the control interval. SCCS has a more flexible definition of time scale and can be defined in terms of calendar time or age, enabling nonparametric approaches to address time-varying confounding. The design can either be restricted to vaccinated individuals or use all cases (vaccinated or unvaccinated). The preselected observation period facilitates evaluations with multiple exposures and recurrent but independent HOIs (26). Compared with the SCRI design, the SCCS design may be more susceptible to bias because of time-varying confounding, as the observation period is often longer than that in the SCRI design. Furthermore, the SCCS design may identify a larger number of cases

requiring more resources than the SCRI design for chart review when the study population is not restricted to vaccinated cases. The SCCS design may be advantageous when identification of a vaccinated group is challenging and the outcome is rare.

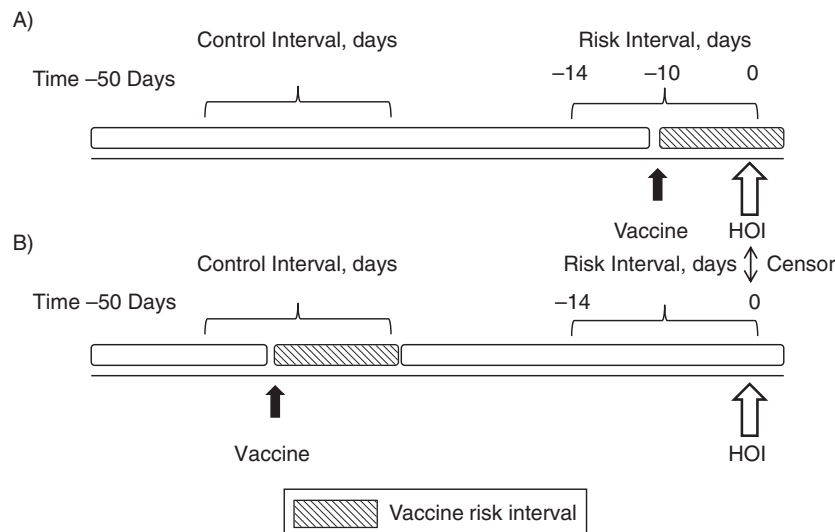
The SCRI design has been used in multiple vaccine-related studies (9, 10, 30–32). Yih et al. (9, 10) used this study design to evaluate the risk of intussusception among infants following rotavirus vaccination in PRISM. This study design was selected to control for time-stable confounders and to avoid exposure misclassification, because only vaccinated individuals were included.

Sun et al. (33) used an SCCS analysis in addition to a cohort analysis to evaluate the risk of febrile seizures after multiple vaccinations with the acellular pertussis vaccine that has been included in combination with diphtheria tetanus toxoids-inactivated poliovirus-*Haemophilus influenzae* type b. The SCCS design facilitated the evaluation of multiple exposures and utilized unexposed time prior to vaccination as well as after vaccination, excluding the 2-week prevaccination period.

#### CASE-CROSSOVER DESIGN AND CASE-TIME-CONTROL DESIGN

The case-crossover design is a modified case-control study comprising only cases where each case contributes both a self-matched case and control window (Figure 6). Similar to the SCRI and SCCS designs, the case-crossover design controls for time-stable confounders, including site heterogeneity, and is susceptible to time-varying confounding. Conditional logistic regression is used to estimate the odds ratio,





**Figure 6.** Case-crossover design. A) Subject 1; B) subject 2. Subjects who experience at least 1 health outcome of interest (HOI) are included, and the design is self-controlled. The case-crossover method is traditionally unidirectional, and individuals are censored at the outcome. In example 1, the individual who experiences the HOI is vaccinated within the risk interval. In example 2, the individual who experiences the HOI is vaccinated in the control interval.

the odds of vaccination in the control window as compared with the time period considered to be the risk window.

Traditionally, the case-crossover design is unidirectional, unlike the SCCS design, and the person-time is censored at the outcome, limiting observation time but potentially reducing the bias due to reverse causality (28, 34). The unidirectional design, however, does not fully eliminate reverse-causality bias that might include bias due to contraindication of the vaccine (28, 35). Furthermore, this design is susceptible to exposure trend bias that may occur, for example, because of a change in policy for a vaccine where all individuals of a certain age are vaccinated and the case window is therefore more heavily exposed than the control window. In order to control for exposure trend bias, a bidirectional case-crossover design or a case-time-control design can be used (28, 36).

The case-time-control design is an extension of the case-crossover study (36). It is a traditional unidirectional case-crossover study with the addition of a time-matched control group. This control group enables adjustment for time trends in exposure. An odds ratio is calculated in both the case group and the noncase group, the odds ratio of the noncase group is an estimate of the period effect because of time trends in exposure, and the ratio of 2 odds ratios is an estimate of the exposure effect (28). The design assumes that the time trend in exposure is comparable in the case and control groups.

PRISM investigators will use a case-time-control design as part of their framework for examining prespecified birth outcomes following vaccination. In the study to evaluate the association between maternal influenza vaccination and the birth outcome of cleft lip or palate (37), cases will be matched to controls on maternal age and estimated date of conception to adjust for time trends in influenza vaccination due to seasonality and gestational age. Among cases, the odds ratio will be

calculated by comparing the odds of vaccination in a risk interval corresponding to the relevant period of organogenesis versus the odds of vaccination in a control interval. A similar odds ratio will be calculated among controls to account for time trends in exposure. Only cases and controls with a vaccination in either the risk or control intervals will be informative for estimating the odds ratio. A marked strength of the case-time-control design in this setting is that restricting to vaccinated individuals avoids confounding by indication (36) and reduces misclassification of exposure to influenza vaccines, which may not be fully captured in the PRISM database.

## CONCLUSION

This overview describes how epidemiologic methods have been tailored to evaluate vaccine safety. Postmarket safety evaluations pose special challenges, because vaccines are often recommended for all persons in a given age group, and those who are not vaccinated may differ systematically in ways that are not easily measured with existing data. Across the study designs, 2 techniques are commonly used to address this issue: data restriction and matching. Data restriction (e.g., case-only or vaccinated case-only) mitigates exposure misclassification and unmeasured differences between vaccinated and unvaccinated persons. Data restriction has pragmatic advantages in terms of protecting data privacy by limiting chart review needs and reducing data collection, review, and management resources. Additionally, matching (e.g., self-matching, time-matched controls, or matching on predefined patient characteristics) is often implemented to address confounding or bias from exposure time trends or other factors. Vaccine safety is particularly sensitive to exposure time trends, because substantial population level shifts can occur whenever

new vaccine use recommendations are issued by the Advisory Committee on Immunization Practices. Such changes may involve completely novel vaccines, newer formulations that expand strain coverage, or the introduction of existing vaccines into new populations (e.g., pregnant women). All methods surveyed here are susceptible to time-varying confounding to different degrees, and many designs require assuming a pre-specified brief interval of elevated risk with a corresponding period of baseline risk, limiting these methods to outcomes with acute onset.

There is no one best study design for all vaccine-outcome pairs. The most important factors in design choice are the characteristics of the HOI (abrupt or insidious onset) and the strength of within- and between-person confounding. For example, in self-controlled analyses, the ability to accurately determine the HOI onset is especially important, because incorrect identification of onset time can result in misattribution of outcomes to the risk or control interval. Self-controlled designs are widely used in vaccine safety, because within-person confounding is minimal and, for many HOIs, the outcome is abrupt, occurs shortly after exposure, and has an onset that can be accurately determined.

Notably, none of the specialized methods we review is inherently superior to a traditional cohort design with regression, and all of them have limitations. In particular, the very data restrictions of self-controlled designs that create resource efficiencies and enable superior confounding control also involve a trade-off in that the risk inferences apply only to vaccinated populations. Thus, case-control and cohort designs that are optimized for distributed databases will always be useful when contrasts are desired to truly unexposed populations. Indeed, even classic case-control designs have been adapted and continue to be valuable to evaluate vaccine safety in pregnancy, because they allow matching and simultaneous adjustment for multiple time trends such as gestational age of vaccination and seasonality. Finally, when applied to vaccine safety using dichotomous outcomes, all 7 methods generate estimates on the relative scale and must be translated to an absolute scale to be most useful to policy makers. This translation often requires baseline rates and introduces a second source of variation that must be accounted for in the final attributable risk estimate. Although the structured decision provides an overview of the study design selection process, it does not include all potential options and implicitly considers only resource costs related to distributed data environments. Moreover, the decision table often presents several design choices in acknowledgement of the complexities of attempting to provide guidance for the myriad of possible vaccine-outcome pairs. Sentinel investigators continue to develop and test innovative methods for vaccine safety surveillance, including strategies for confounder adjustment and analyses in distributed data settings to complement the methods described in this framework.

## ACKNOWLEDGMENTS

Author affiliations: Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care

*Am J Epidemiol.* 2015;181(8):608–618

Institute, Boston, Massachusetts (Meghan A. Baker, Tracy A. Lieu, Lingling Li, Alison Tse Kawai); Division of Infectious Diseases, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts (Meghan A. Baker); Division of Research, Kaiser Permanente Northern California, Oakland, California (Tracy A. Lieu, Bruce H. Fireman); and Center for Biologics Evaluation and Research, US Food and Drug Administration, Silver Spring, Maryland (Wei Hua, Yandong Qiang, David B. Martin, Michael D. Nguyen).

This work was supported by the Food and Drug Administration through the Department of Health and Human Services (contract HHSF2232009100061).

We thank Dr. Richard Platt for his input and for his support via the leadership of the Mini-Sentinel Program and its distributed database. We greatly appreciate the contributions of Dr. Grace M. Lee, Dr. W. Katherine Yih, Dr. Joshua J. Gagne, Dr. Sharon K. Greene, and the PRISM team on the study design selection framework. We also thank Ashleigh Goff for her help with figures and references.

Conflict of interest: none declared.

## REFERENCES

1. Nguyen M, Ball R, Midthun K, et al. The Food and Drug Administration's Post-Licensure Rapid Immunization Safety Monitoring program: strengthening the federal vaccine safety enterprise. *Pharmacoepidemiol Drug Saf.* 2012;21(suppl 1): 291–297.
2. Curtis LH, Weiner MG, Boudreau DM, et al. Design considerations, architecture, and use of the Mini-Sentinel distributed data system. *Pharmacoepidemiol Drug Saf.* 2012; 21(suppl 1):23–31.
3. Gagne JJ, Fireman B, Ryan P, et al. Taxonomy for monitoring methods within a medical product safety surveillance system: report of the Mini-Sentinel Taxonomy Project Work Group. [http://www.mini-sentinel.org/work\\_products/Statistical\\_Methods/Mini-Sentinel\\_FinalTaxonomyReport.pdf](http://www.mini-sentinel.org/work_products/Statistical_Methods/Mini-Sentinel_FinalTaxonomyReport.pdf). Published September 22, 2010. Accessed September 10, 2012.
4. Gagne JJ, Nelson JC, Fireman B, et al. Taxonomy for monitoring methods within a medical product safety surveillance system: year two report of the Mini-Sentinel Taxonomy Project Workgroup. [http://www.mini-sentinel.org/work\\_products/Statistical\\_Methods/Mini-Sentinel\\_Methods\\_Taxonomy-Year-2-Report.pdf](http://www.mini-sentinel.org/work_products/Statistical_Methods/Mini-Sentinel_Methods_Taxonomy-Year-2-Report.pdf). Published January 31, 2012. Accessed September 10, 2012.
5. Gagne JJ, Fireman B, Ryan PB, et al. Design considerations in an active medical product safety monitoring system. *Pharmacoepidemiol Drug Saf.* 2012;21(suppl 1):32–40.
6. Glanz JM, McClure DL, Xu S, et al. Four different study designs to evaluate vaccine safety were equally validated with contrasting limitations. *J Clin Epidemiol.* 2006;59(8):808–818.
7. Ryan PB, Madigan D, Stang PE, et al. Empirical assessment of methods for risk identification in healthcare data: results from the experiments of the Observational Medical Outcomes Partnership. *Stat Med.* 2012;31(30):4401–4415.
8. Kawai AT, Martin DB, Kulldorf M, et al. Assessment of febrile seizures after trivalent influenza vaccines during the 2010–2011 influenza season in Post-licensure Rapid Immunization Safety Monitoring Program. [http://www.mini-sentinel.org/work\\_products/PRISM/Mini-Sentinel\\_PRISM\\_Influenza-Vaccines-and-Febrile-Seizures-Report.pdf](http://www.mini-sentinel.org/work_products/PRISM/Mini-Sentinel_PRISM_Influenza-Vaccines-and-Febrile-Seizures-Report.pdf). Published May 15, 2014. Accessed May 20, 2014.

9. Yih WK, Lieu TA, Kulldorff M, et al. Intussusception risk after rotavirus vaccination in U.S. infants. *N Engl J Med*. 2014; 370(6):503–512.
10. Yih WK, Lieu TA, Kulldorff M, et al. Intussusception risk after rotavirus vaccination in U.S. infants. [http://www.mini-sentinel.org/work\\_products/PRISM/Mini-Sentinel\\_PRISM\\_Rotavirus-and-Intussusception-Report.pdf](http://www.mini-sentinel.org/work_products/PRISM/Mini-Sentinel_PRISM_Rotavirus-and-Intussusception-Report.pdf). Published June 12, 2013. Accessed June 14, 2013.
11. Tse A, Tseng HF, Greene SK, et al. Signal identification and evaluation for risk of febrile seizures in children following trivalent inactivated influenza vaccine in the Vaccine Safety Datalink Project, 2010–2011. *Vaccine*. 2012;30(11):2024–2031.
12. Shui IM, Baggs J, Patel M, et al. Risk of intussusception following administration of a pentavalent rotavirus vaccine in US infants. *JAMA*. 2012;307(6):598–604.
13. Gagne JJ, Baker M, Bykov K, et al. Development of the Mini-Sentinel taxonomy prompt selection tool: year three report of the Mini-Sentinel Taxonomy Project Workgroup. [http://www.mini-sentinel.org/work\\_products/Statistical\\_Methods/Mini-Sentinel\\_Methods\\_Taxonomy-Year-3-Report.pdf](http://www.mini-sentinel.org/work_products/Statistical_Methods/Mini-Sentinel_Methods_Taxonomy-Year-3-Report.pdf). Published September 30, 2013. Accessed May 18, 2014.
14. Rothman KJ, Greenland S. *Modern Epidemiology*. 2nd ed. Philadelphia, PA: Lippincott-Raven; 1998.
15. Szklo M, Nieto FJ. *Epidemiology: Beyond the Basics*. 2nd ed. Sudbury, MA: Jones and Bartlett Publishers; 2007.
16. Fine PE, Chen RT. Confounding in studies of adverse reactions to vaccines. *Am J Epidemiol*. 1992;136(2):121–135.
17. Jackson LA, Nelson JC, Benson P, et al. Functional status is a confounder of the association of influenza vaccine and risk of all cause mortality in seniors. *Int J Epidemiol*. 2006;35(2):345–352.
18. Dublin S, Walker R, Rutter C, et al. Collecting supplemental information via two-phase study designs to investigate signals arising from medication safety surveillance activities. [http://www.mini-sentinel.org/work\\_products/Statistical\\_Methods/Mini-Sentinel\\_Methods\\_Supplemental-Information\\_Two-Phase-Study-Designs.pdf](http://www.mini-sentinel.org/work_products/Statistical_Methods/Mini-Sentinel_Methods_Supplemental-Information_Two-Phase-Study-Designs.pdf). Published December 18, 2013. Accessed May 18, 2014.
19. Fireman B, Lee J, Lewis N, et al. Influenza vaccination and mortality: differentiating vaccine effects from bias. *Am J Epidemiol*. 2009;170(5):650–656.
20. Rowhani-Rahbar A, Klein NP, Lewis N, et al. Immunization and Bell's palsy in children: a case-centered analysis. *Am J Epidemiol*. 2012;175(9):878–885.
21. France EK, Glanz J, Xu S, et al. Risk of immune thrombocytopenic purpura after measles-mumps-rubella immunization in children. *Pediatrics*. 2008;121(3):e687–e692.
22. Murphy TV, Gargiullo PM, Massoudi MS, et al. Intussusception among infants given an oral rotavirus vaccine. *N Engl J Med*. 2001;344(8):564–572.
23. Black S, Shinefield H, Ray P, et al. Risk of hospitalization because of aseptic meningitis after measles-mumps-rubella vaccination in one- to two-year-old children: an analysis of the Vaccine Safety Datalink (VSD) Project. *Pediatr Infect Dis J*. 1997;16(5):500–503.
24. Irving SA, Kieke BA, Donahue JG, et al. Trivalent inactivated influenza vaccine and spontaneous abortion. *Obstet Gynecol*. 2013;121(1):159–165.
25. Farrington CP. Relative incidence estimation from case series for vaccine safety evaluation. *Biometrics*. 1995;51(1):228–235.
26. Whitaker HJ, Farrington CP, Spiessens B, et al. Tutorial in biostatistics: the self-controlled case series method. *Stat Med*. 2006;25(10):1768–1797.
27. Farrington CP, Whitaker HJ, Hocene MN. Case series analysis for censored, perturbed, or curtailed post-event exposures. *Biostatistics*. 2009;10(1):3–16.
28. Madigan D, Fireman B, Maclure M. Case-based methods workgroup report. Section 1. [http://www.mini-sentinel.org/work\\_products/Statistical\\_Methods/Mini-Sentinel\\_Methods\\_Case-Based-Report.pdf](http://www.mini-sentinel.org/work_products/Statistical_Methods/Mini-Sentinel_Methods_Case-Based-Report.pdf). Published October 4, 2011. Accessed February 4, 2012.
29. Weldeselasie YG, Whitaker HJ, Farrington CP. Use of the self-controlled case-series method in vaccine safety studies: review and recommendations for best practice. *Epidemiol Infect*. 2011;139(12):1805–1817.
30. France EK, Glanz JM, Xu S, et al. Safety of the trivalent inactivated influenza vaccine among children: a population-based study. *Arch Pediatr Adolesc Med*. 2004; 158(11):1031–1036.
31. Kramarz P, DeStefano F, Gargiullo PM, et al. Does influenza vaccination exacerbate asthma? Analysis of a large cohort of children with asthma. Vaccine Safety Datalink Team. *Arch Fam Med*. 2000;9(7):617–623.
32. O'Leary ST, Glanz JM, McClure DL, et al. The risk of immune thrombocytopenic purpura after vaccination in children and adolescents. *Pediatrics*. 2012;129(2):248–255.
33. Sun Y, Christensen J, Hviid A, et al. Risk of febrile seizures and epilepsy after vaccination with diphtheria, tetanus, acellular pertussis, inactivated poliovirus, and *Haemophilus influenzae* type B. *JAMA*. 2012;307(8):823–831.
34. Maclure M. The case-crossover design: a method for studying transient effects on the risk of acute events. *Am J Epidemiol*. 1991;133(2):144–153.
35. Salas M, Hofman A, Stricker BH. Confounding by indication: an example of variation in the use of epidemiologic terminology. *Am J Epidemiol*. 1999;149(11):981–983.
36. Suissa S. The case-time-control design. *Epidemiology*. 1995; 6(3):248–253.
37. Kawai A, Li L, Andrade S, et al. Influenza vaccines and birth outcomes. [http://www.mini-sentinel.org/work\\_products/PRISM/Mini-Sentinel\\_PRISM\\_Influenza-Vaccines-and-Birth-Outcomes-Protocol.pdf](http://www.mini-sentinel.org/work_products/PRISM/Mini-Sentinel_PRISM_Influenza-Vaccines-and-Birth-Outcomes-Protocol.pdf). Published February 25, 2013. Updated February 26, 2013. Accessed May 18, 2014.